

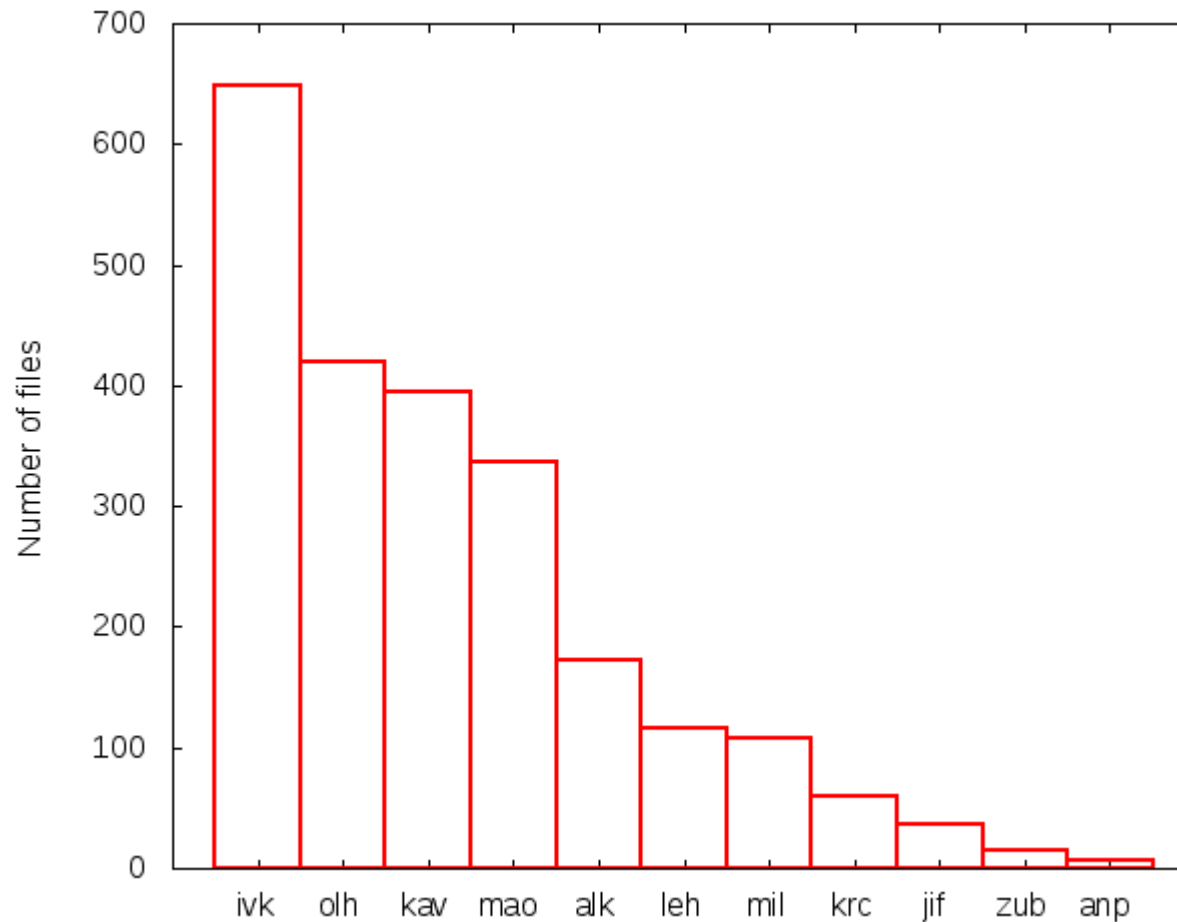
# Czech a-layer

- Only the files revised at the time of annotation (2000 sentences, June – July 2008)
- 1 annotator, 1 person to check the annotation
- Sections 22 & 23, used in CoNLL-2009 ST (OOD data)

2200	2201	2202	2203	2204	2205	2206	2207	2208	2209
2210	2211	2212	2213	2214	2215	2216	2217	2218	2219
2220	2221	2222	2223	2224	2225	2226	2227	2228	2229
2230	2231	2232	2233	2234	2235	2236	2237	2238	2239
2240	2241	2242	2243	2244	2245	2246	2247	2248	2249
2250	2251	2252	2253	2254	2255	2256	2257	2258	2259
2260	2261	2262	2263	2264	2265	2266	2267	2268	2269
2270	2271	2272	2273	2274	2275	2276	2277	2278	2279
2280	2281	2282							
2300	2301	2302	2303	2304	2305	2306	2307	2308	2309
2310	2311	2312	2313	2314	2315	2316	2317	2318	2319
2320	2321	2322	2323	2324	2325	2326	2327	2328	2329
2330	2331	2332	2333	2334	2335	2336	2337	2338	2339
2340	2341	2342	2343	2344	2345	2346	2347	2348	2349
2350	2351	2352	2353	2354	2355	2356	2357	2358	2359
2360	2361	2362	2363	2364	2365	2366	2367	2368	2369
2370	2371	2372	2373	2374	2375	2376	2377	2378	2379
2380	2381	2382	2383	2384	2385	2386	2387	2388	2389
2390	2391	2392	2393	2394	2395	2396	2397	2398	

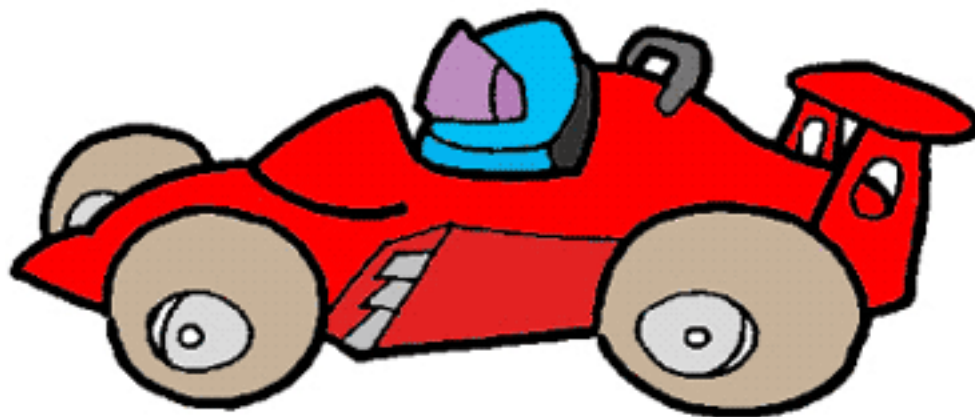
# Annotation of Czech

- 11 annotators (max. 8 at the same time), 1 more person editing vallex



# Annotation of Czech (2)

- Annotation speed: 9.2 sentences per hour (49.000 sentences → 5.326 hours)



- Parallel annotation only to measure inter-annotator agreement (62 files, 155 instances)
- Automatic data checking (147 error types)

# Comparison en/cs

	English	Czech
Number of tokens	1,234,200	1,151,200
Number of t-nodes	838,200	932,700

# Comparison PCEDT/PDT

- Manually:
  - structure
  - functors
  - valency
  - links to a-layer
  - coreference
- Missing:
  - subfunctors
  - topic-focus articulation
  - quotation types
  - name of person
- Automatic:
  - nodetype
  - grammatememes

# Comparison PCEDT/PDT (2)

- Valency
  - Verbs: full information
  - Deverbative nouns (*-ní, -tí, e.g. balení*)
    - Denoting an event: valency, missing obligatory elements plus coreference, but not the lexicon
  - Other nouns
    - “Everything” is RSTR (no actants, MAT, APP)
  - Other POS
    - Actants only for surface elements

# Comparison PCEDT/PDT (3)

- No TFA: the position of RHEM not reliable
- Structure:simplification
  - “red and white wine” not distinguishable from “red and white flag”

# Comparison PCEDT/PDT (4)

- Coreference
  - T-lemma of added nodes (#NewNode) determined by annotators of coreference
  - No #QCor

